The Málaga Urban Dataset: High-rate Stereo and Lidars in a realistic urban scenario

José-Luis Blanco-Claraco*, Francisco-Ángel Moreno-Dueñas** and Javier González-Jiménez**

*Department of Mechanical Engineering, University of Almería
**Department of System Engineering and Automation, University of Málaga

September 13, 2018

Abstract

This paper introduces a dataset gathered entirely in urban scenarios with a car equipped with one stereo camera and five laser scanners, among other sensors. One distinctive feature of the present dataset is the existence of high-resolution stereo images grabbed at high rate (20 fps) during a 36.8 km trajectory, which allows the benchmarking of a variety of computer vision techniques. We describe the employed sensors and highlight some applications which could be benchmarked with the presented work. Both plain text and binary files are provided, as well as open source tools for working with the binary versions. The dataset is available for download in `http://www.mrpt.org/MalagaUrbanDataset`.

1 Introduction

Applying the scientific method to computer vision and Simultaneous Localization and Mapping (SLAM) implies being able to perform rigorous benchmarking of the different algorithms in order to determine their suitability and relative performance.

The interest of the community in this sense is clear, given the number of projects and workshops devoted to the topic [Bonarini et al., 2006, Sturm et al., 2012]. One of the best known SLAM dataset is the *Sydney Victoria park* dataset [Guivant and Nebot, 2001], which is however limited to 2D range-bearing mapping. The project Rawseeds [Bonarini et al., 2006] also aimed at providing indoor and outdoor datasets with visual information. More recent releases that include images of urban areas [Geiger et al., 2012, Peynot et al., 2010] or both images and laser data of park-like zones [Smith et al., 2009] have also received the attention of the community, clearly reflecting the demand for this kind of releases.

In comparison to previous datasets, and as summarized with Table 1, we claim that the present work provides a unique combination of (i) multiple laser scanners pointing in various orientations and (ii) high-rate (20 fps) and high-resolution (1024×768) *stereo images* of good quality (e.g. minimal motion blur). In addition, a significant part of our dataset reflects dynamic environments with real-life traffic, thus becoming a challenging testbed for SLAM, visual odometry and object detection methods.

The structure of this paper is as follows. Section 2 addresses the configuration of the vehicle, next we describe each employed sensor and finally section 3 presents the dataset itself.

Dataset	GPS	GT	IMU	Laser scanners	Images	Path length
New College (2008) and Oxford city center [Cummins and Newman, 2008]	✓	×	×	×	Mono.: color 640 × 480@ ∼ 1 fps	College: 2 km City: 28 km
New College (2009) [Smith et al., 2009]	✓	×	✓	2	Stereo: b/w 512 × 384@20 fps Ladybug: 5 × (color 384 × 512@5 fps)	2.2 km
Rawseeds datasets (2009) [Ceriani et al., 2009]	✓	✓	✓	4	Front: color 320 × 240@29.95 fps Omni.: b/w 640 × 640@15 fps Trino.: b/w 640 × 480@15 fps	Indoors: 0.89 km Outdoors: 1.9 km
Málaga 2009 dataset [Blanco et al., 2009]	✓	✓	✓	5	Stereo: color 1024 × 768@7.5 fps	6 km
MIT DARPA [Huang et al., 2010]	✓	✓	✓	13	Mono.: 4 × (color 376 × 240@10 fps) Mono.: color 752 × 480@22.8 fps	90 km
The Marulan datasets [Peynot et al., 2010]	✓	✓	✓	4	Mono.: b/w 1360 × 1024@10 fps	∼ 1 km
Karlsruhe sequences [Geiger et al., 2010] [Geiger et al., 2011]	✓	✓	✓	×	Stereo: b/w 1344 × 391@10 fps	6.9 km
Ford campus [Pandey et al., 2011]	✓	✓	✓	1	Omni.: color 1600 × 600@8 fps	∼ 6 km
KITTI [Geiger et al., 2012]	✓	✓	✓	1	Stereo: b/w 1392 × 512@10 fps Stereo: color 1392 × 512@10 fps	∼ 50 km
Málaga Urban dataset (this work)	✓	×	✓	5	Stereo: color 1024 × 768@20 fps	36.8 km

Table 1: A comparison of some previous datasets regarding the presence (✓) or not (×) of GPS sensors, ground truth (GT), inertial units (IMU), the usage of laser scaners, the kind of cameras on the vehicle and the dataset path lengths.

2 Vehicle setup

2.1 Physical characteristics

In order to be able to navigate outdoors in a safe way throughout typical urban scenarios we decided to employ a common *Citroen C4* car, shown in Figure 1. All the sensors were installed in a modified roof-rack, designed for a flexible placement of heterogeneous devices. This configuration allows us to drive among the city traffic without restrictions.

Two computers were also installed inside the vehicle to cope with the computational and storage bandwidth requirements. All the electrical power for computers and sensors was obtained from the vehicle's own power system.

We recorded data from an overall of eight sensors: one stereo camera, five laser scanners, one inertial measurement unit (IMU) and one GPS receiver.

Figure 2 schematically illustrates the placement of each sensor on the vehicle, with approximate (hand-measured) coordinates shown in Table 2. The local frame of reference is set such that the positive x axis always points forwards and z points upwards, as customary in mobile robotics.

Next we briefly describe the relevant characteristics of each sensor and the reasons for their inclusion in the dataset.

2.1.1 Stereo camera

Color imaging was provided by a Point Grey Research's *Bumblebee 2* stereo camera, configured to capture images at its maximum resolution of 1024 × 768 at 20fps. As opposed to our previous dataset [Blanco et al., 2009], the usage of a stereo camera instead of two independent ones assures a precise synchronization in both image streams. The camera gain and white-balance control were left in automatic mode.

Once one determines the camera intrinsic parameters, the rigid mounting of the two CCD sensors inside the camera and the use of a fixed focal distance lead to a reliable calibration that is not affected by

Sensor	x (m)	y (m)	z (m)	***yaw*** (deg.)	***pitch*** (deg.)	***roll*** (deg.)
`CAMERA1`	0.785	0	~~0.273~~ 0.057	0	−8.2	0
`XSensMTi`	0.400	0.040	0.000	0	0	0
`GPS_DELUO`	0.155	0.069	0.004	n/a	n/a	n/a
`LASER1` (Rear SICK)	−0.023	0	0.097	−180	0	0
`LASER2` (Front SICK)	0.536	0	0.093	0	0	0
`HOKUYO1` (Front)	0.536	0	0.273	0	21.4	0
`HOKUYO2` (Right)	0.075	−0.489	0.055	−90	0	−90
`HOKUYO3` (Left)	0.075	0.489	0.055	90	0	90

(n/a: not applicable)

Table 2: Summary of approximate sensor positioning on the vehicle. Refer to Figure 2.

(a)

(b)

Figure 1: The instrumented vehicle employed for collecting the dataset: (a) general view and (b) close-up of the sensors.

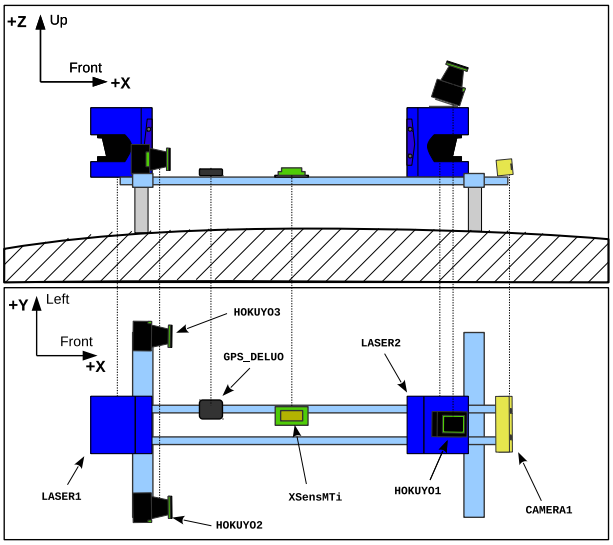


Figure 2: Side and top views, respectively, of the relative positions of sensors on the vehicle's roof-rack structure. Compare to Figure 1(b). Not to scale.

shocks and vibrations. Although the dataset includes all camera calibration parameters, we also publish a collection of raw stereo images of a checkerboard to allow the reader applying different calibration methods.

This camera was placed pointing forwards and slightly tilted up, to avoid capturing part of the vehicle chassis. Despite the small parallax obtained during navigation from forward-looking cameras, we decided to use this configuration for its interesting applications in detecting other vehicles, pedestrians, traffic lights, etc.

2.1.2 Laser scanners

The vehicle was equipped with five laser scanners: three Hokuyo UTM-30LX and two SICK LMS-200. The former are small, energy-efficient scanners with a range of 30 meters and a field of view of 270°. With an angular resolution of 0.25°, they provide nominal accuracies of 30mm and 50mm for distances below and above 10 meters, respectively. The latter models, manufactured by SICK for industrial use, are considerably heavier, more robust and more energy demanding. In turn, their working range extends up to 80 meters and are less prone to detecting "phantom points" near sharp edges, a problem occasionally found in range data from the Hokuyo sensor.

Regarding the placement of the scanners onboard, they can be divided in three groups:

- The two SICK scanners sense in the horizontal plane. These data may be useful for 2D SLAM for parts of the trajectory that are flat enough.
- Two lateral Hokuyo sensors provide a vertical scanning of the vehicle surroundings.
- Finally, one Hokuyo scanner is placed pointing forwards and tilted down, in order to sequentially scan the road ahead the vehicle, e.g. for detecting obstacles.

The two scanners (SICK and Hokuyo) pointing forwards may find applications to detection algorithms that fuse visual and range information.

2.1.3 IMU

Inertial sensors based on inexpensive MEMS technology are present nowadays in many portable devices such as tablets or smartphones. Therefore, it seems reasonable to try to explore the possibilities that these sensors create for improving visual odometry or visual SLAM methods.

To endow our dataset with this kind of information we installed an *xSens MTi* inertial unit on the vehicle. It was firmly attached to the roof structure like all the other sensors, thus the angular velocities sensed by the device can be accurately assigned to the rest of sensors as well, disregarding the negligible effects of the structure elastic deformations during the drive.

With a rate of 100 Hz, the measurements provided by this device include:

- 3-axis acceleration. We have experimentally measured its static error, which has a standard deviation of $\sigma_{acc} \approx 0.05 m/s^2$.
- 3-axis instantaneous angular velocity. Its experimental angular velocity error has been found to be $\sigma_{gyro} \approx 0.4°/s$, while systematic errors were noticed for *yaw* (rotations around the Z axis) in the order of $\sim 0.6°/s$.
- Attitude dead-reckoning in 3D, as provided by the internal filter implemented by the manufacturer.

2.1.4 GPS receiver

We also installed a consumer-grade, low-cost GPS receiver on the car, with a two-fold purpose: (i) providing approximate positioning for a better understanding of the whole trajectory traversed in this dataset (see Figure 3), and (ii) offering realistic GPS data for usage in visual SLAM applications aimed at the automotive industry.

This sensor provides positioning data at 1 Hz during the whole dataset, with the exception of a few unavoidable segments ("urban canyons" and dense groves) were the signal was too weak to provide good localization.

Two additional industry-grade GPS receivers were also installed in the vehicle (*mmGPS* devices from

Figure 3: An overview of the complete trajectory, as reconstructed from GPS data. A zoomable version is available online.

Topcon, the two cylindrical yellow devices in Figure 1), but positioning information was not available from these receivers during the recording of the dataset. However, frames with GPS timing information were collected from both receivers in order to accurately synchronize the local clocks of the two computers. By grabbing satellite timestamps from two identical receivers in both computers we have been able to establish a least-square fit of the mapping between the reference GPS time and the local clocks. More importantly, this mapping provides an accurate way of merging the partial datasets grabbed in each machine during an offline postprocessing stage. Interestingly, we found out that not only the local clocks had an offset (as could be expected) but that they exhibit a small drift (6.06 $\mu s/s$ and 83.33 $\mu s/s$, respectively), which has been corrected in the published dataset.

2.2 Software

The vehicle is equipped with sensors of quite different types, each generating data at different rates. Thus, the software intended to record the data logs must be capable of dealing with asynchronous streams from the sensors. For this purpose, we employed the data logger application *rawlog-grabber*, as we also did for previous datasets [Blanco et al., 2009].

This program launches one thread for each individual sensor. Then, each thread splits the sensory data into their corresponding natural discrete pieces (called *observations*), e.g. a complete 2D scan for laser scanners, and marks them with timestamps. Since our system does not run on a real-time OS, we have to assure that no observation is lost by creating a FIFO queue for each thread, then merging all of their outputs into a thread-safe timestamp-sorted queue, which is periodically pushed to a binary *"rawlog"* file. We chose binary log files for their bandwidth efficiency in contrast to other pure-text formats. Afterwards, we have post-processed the binary logs to generate plain text logs for the convenience of readers.

Collecting large images (1024×768) at real-time without dropping frames presented an additional challenge, because hard-disk bandwidth is not enough for saving raw images, while lossy compression solves the issue but introduces a high computa-

Sensor label	**Count**	**Duration** (sec)	**Actual rate** (Hz)	**Nominal rate** (Hz)
CAMERA1	113082	5654.6	19.998	20
GPS_DELUO	11244	5653.0	1.989	1
HOKUYO1	225416	5654.62	39.864	40
HOKUYO2	225631	5654.62	39.902	40
HOKUYO3	225510	5654.62	39.880	40
LASER1	398531	5315.58	74.974	75
LASER2	404487	5498.11	73.568	75
XSensMTi	549816	5498.15	100.000	100

Table 3: Summary of grabbed data from each sensor. The *actual rates* shown here are the average values obtained as the ratio *count/ duration*.

tional burden. Our approach consisted in parallelizing the latter task by creating additional threads with the sole purpose of compressing images in a high-quality format (JPEG format, quality=95).

3 Dataset summary

The following paragraphs describe the most relevant characteristics and statistics about the presented dataset. However, accessing to the supplementary material online[1] is recommended for having a better insight about its content.

3.1 Description

The dataset was recorded as a single sequence during a car trip throughout different urban areas of Málaga, with a total duration of ∼93 minutes. An overlaid impression of the GPS-reconstructed path over a map of the city is provided in Figure 3.

Observations from all sensors were recorded at their maximum nominal rates. These values, along with the actual average rates obtained from the logged stream of data, are shown in Table 3. The similarity of actual and nominal rates means that only a tiny fraction of sensory data was dropped for most sensors (mostly due to corrupt frames for communication errors), with the worst case being the sensor LASER2 (front SICK laser) for which a 1.9% of all frames were lost. An overall of 2.2 millions of individual *observations* were collected.

Regarding the trajectory followed during the recording, we can split the dataset into the following *segments* or *epochs* (within parentheses, the starting and end points measured in minutes since start):

- Epoch 1 (0–6min): Four loops within the parking lot of the Computer Science School of the University of Málaga. This area was also recorded 13 months earlier for a previous dataset with a different camera [Blanco et al., 2009], making this segment ideal for testing place recognition algorithms.
- Epoch 2 (6–10min): Driving towards a nearby suburb, crossing one under-construction road.
- Epoch 3 (10–52min): One of the main parts of the dataset, in which North-West Málaga suburbs ("El Cónsul" and "El Romeral") are transversed several times including nested loop closures. The car underwent a parking maneuvering during minutes 17–19. Traffic lights and take overs also appear in this segment.
- Epoch 4 (52–60min): A trip towards downtown, traversing a highway-like road. In contrast to the velocity range of 20–40 km/h (12.4–24.9 mph) in the other epochs, in this segment the vehicle moves faster than 50 km/h (31 mph).
- Epoch 5 (60–93min): Another of the most interesting segments, since it includes several loop

[1] See: http://www.mrpt.org/MalagaUrbanDataset.

closures in downtown. Here we find the highest traffic density for the entire dataset.

As an additional tool to help the interested reader to pick relevant segments from the dataset we created a *video index* (see Figure 5), available online[2]. Apart from camera images, the video shows a 3D point cloud reconstruction of the environment from the vertical laser scanners and GPS data as a gross estimate of the ground truth path. Some snapshots of the obtained scenarios can be also seen in Figure 4.

In order to make working with the dataset easier, it has been further divided into 15 smaller sequences or *extracts*, illustrated in Figure 6. A video is also available online for each individual sequence, such that they can be easily inspected. Next, we enumerate the length in seconds of each extract and provide a brief description of its contents:

1. Straight path in the faculty parking (39 s).
2. Through an under-construction road (92 s).
3. Three-quarters of a turn in a roundabout (41 s).
4. Crossing a roundabout, some traffic (32 s).
5. Loop closure ($\sim$ 1.7 km) in a straight avenue (240 s).
6. Loop closure ($\sim$ 1.2 km) around building blocks (230 s).
7. Loop closure ($\sim$ 0.7 km) around a small avenue (106 s).
8. Long loop closure ($\sim$ 4.5 km) (501 s).
9. Through the campus boulevard, with some traffic (50 s).
10. Multiple loop closures in a suburb area (865 s).
11. High-way incorporation, some traffic (144 s).
12. Long avenue ($\sim$ 3.7 km), dense traffic (443 s).
13. At downtown. Dense traffic and pedestrians (1572 s).
14. Direct sun conditions at a parking area (112 s).
15. Direct sun conditions at a suburb area (69 s).

[2] `http://www.youtube.com/watch?v=tM5BSLKUSxU`

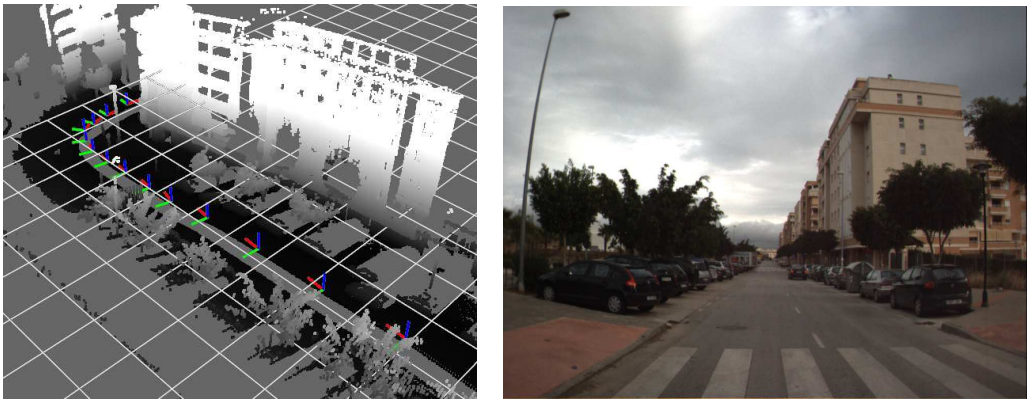
Although all sensory data are provided in plain-text format, it is worth mentioning that two ready-to-use applications (named `RawLogViewer` and `rawlog-edit`) are provided to inspect, filter or split binary log files. These programs are already shipped within modern Debian and Ubuntu GNU/Linux distributions as part of the package `mrpt-apps`. Example C++ source code is also available online for readers interested in parsing binary logs.

3.2 Challenges

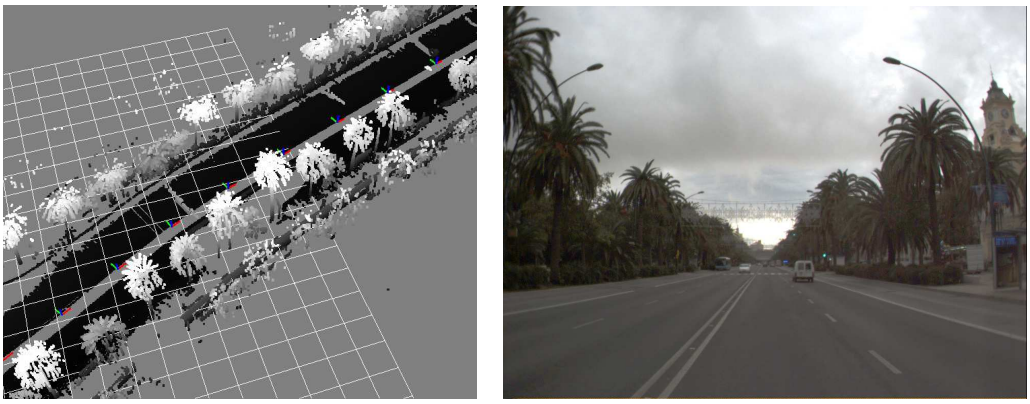
We found that a particularly challenging problem during the recording of outdoor images was the appearance of vertical smears caused by direct sun exposure. After several attempts at different dates we obtained, in a cloudy day, the present dataset which exhibits a minor occurrence of such smears. Another challenging aspect of the images, from the point of view of computer vision, is the dynamic gain control of the camera which may introduce hurdles to feature tracking algorithms. Anyway, we believe that these challenges are intrinsic and unavoidable for any real-world problem where cameras are to be placed on vehicles for navigation in uncontrolled, outdoor scenarios. In order to allow researchers to easily compare diverse robust techniques against this kind of problems, we released two short dataset extracts (numbered #14 and #15) with direct sun exposure.

4 Conclusions

We have presented a dataset whose most relevant component is the presence of high-rate and high-resolution stereo video in unmodified urban scenarios. The authors believe that the mobile robotics community will find it specially suited for benchmarking of visual odometry, visual SLAM and appearance-based recognition methods. Moreover, the presence of several laser scanners enables Lidar-vision object detection and recognition within realistic traffic situations.



(a) t=1278s



(b) t=3732s

(c) t=3983s

Figure 4: Three sample screenshots from the dataset: (left) 3D reconstructions from vertical laser scanners and GPS-only information, (right) images from the stereo camera in the same places.

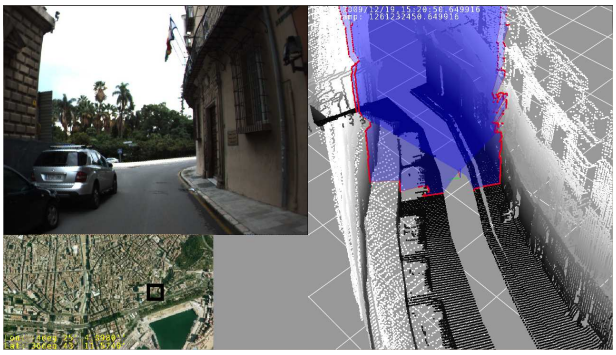


Figure 5: A view of the dataset *video index*, which simultaneously displays: (top-left) raw video frames, (bottom-left) the current location of the vehicle over the city map and (right) local 3D point cloud from laser scanners.

Acknowledgements

This work has been partially funded by the Spanish "Ministerio de Ciencia e Innovación" under contract number DPI2008-03527 and the grant program JDC-MICINN 2011.

References

[Blanco et al., 2009] Blanco, J.-L., Moreno, F.-A., and González, J. (2009). A collection of outdoor robotic datasets with centimeter-accuracy ground truth. *Autonomous Robots*, 27(4):327–351.

[Bonarini et al., 2006] Bonarini, A., Burgard, W., Fontana, G., Matteucci, M., Sorrenti, D. G., and Tardos, J. D. (2006). Rawseeds: Robotics advancement through web-publishing of sensorial and elaborated extensive data sets. In *Proceedings of IROS'06 Workshop on Benchmarks in Robotics Research*.

[Ceriani et al., 2009] Ceriani, S., Fontana, G., Giusti, A., Marzorati, D., Matteucci, M., Migliore, D., Rizzi, D., Sorrenti, D. G., and Taddei, P. (2009). Rawseeds ground truth collection systems for indoor self-localization and mapping. *Autonomous Robots*, 27(4):353–371.

[Cummins and Newman, 2008] Cummins, M. and Newman, P. (2008). FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. *The International Journal of Robotics Research*, 27(6):647–665.

[Geiger et al., 2012] Geiger, A., Lenz, P., and Urtasun, R. (2012). Are we ready for autonomous driving? the kitti vision benchmark suite. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3354–3361.

[Geiger et al., 2010] Geiger, A., Roser, M., and Urtasun, R. (2010). Efficient large-scale stereo matching. In *Asian Conference on Computer Vision (ACCV)*.

[Geiger et al., 2011] Geiger, A., Ziegler, J., and Stiller, C. (2011). Stereoscan: Dense 3d reconstruction in real-time. In *Intelligent Vehicles Symposium (IV)*.

[Guivant and Nebot, 2001] Guivant, J. E. and Nebot, E. M. (2001). Optimization of the simultaneous localization and map-building algorithm for real-time implementation. *IEEE Transactions on Robotics and Automation*, 17(3):242–257.

[Huang et al., 2010] Huang, A. S., Antone, M., Olson, E., Fletcher, L., Moore, D., Teller, S., and Leonard, J. (2010). A high-rate, heterogeneous data set from the darpa urban challenge. *The International Journal of Robotics Research*, 29(13):1595–1601.

[Pandey et al., 2011] Pandey, G., McBride, J. R., and Eustice, R. M. (2011). Ford campus vision and lidar data set. *The International Journal of Robotics Research*, 30(13):1543–1552.

[Peynot et al., 2010] Peynot, T., Scheding, S., and Terho, S. (2010). The Marulan Data Sets: Multi-sensor Perception in a Natural Environment with Challenging Conditions. *The International Journal of Robotics Research*, 29(13):1602.

[Smith et al., 2009] Smith, M., Baldwin, I., Churchill, W., Paul, R., and Newman, P. (2009). The new college vision and laser data set.

Figure 6: Summary of the 15 dataset extracts available for download. For each segment, the vehicle path is shown together with aerial urban images for reference. Refer to the online material for color images and interactive maps.

The International Journal of Robotics Research, 28(5):595.

[Sturm et al., 2012] Sturm, J., Engelhard, N., Endres, F., Burgard, W., and Cremers, D. (2012). A benchmark for the evaluation of rgb-d slam systems. In *IEEE International Conference on Intelligent Robot Systems (IROS)*.